

# Two-part models with stochastic processes for modelling longitudinal semicontinuous data: Computationally efficient inference and modelling the overall marginal mean

Sean Yiu and Brian DM Tom

Statistical Methods in Medical Research

0(0) 1–17

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0962280217710573

journals.sagepub.com/home/smm



## Abstract

Several researchers have described two-part models with patient-specific stochastic processes for analysing longitudinal semicontinuous data. In theory, such models can offer greater flexibility than the standard two-part model with patient-specific random effects. However, in practice, the high dimensional integrations involved in the marginal likelihood (i.e. integrated over the stochastic processes) significantly complicates model fitting. Thus, non-standard computationally intensive procedures based on simulating the marginal likelihood have so far only been proposed. In this paper, we describe an efficient method of implementation by demonstrating how the high dimensional integrations involved in the marginal likelihood can be computed efficiently. Specifically, by using a property of the multivariate normal distribution and the standard marginal cumulative distribution function identity, we transform the marginal likelihood so that the high dimensional integrations are contained in the cumulative distribution function of a multivariate normal distribution, which can then be efficiently evaluated. Hence, maximum likelihood estimation can be used to obtain parameter estimates and asymptotic standard errors (from the observed information matrix) of model parameters. We describe our proposed efficient implementation procedure for the standard two-part model parameterisation and when it is of interest to directly model the overall marginal mean. The methodology is applied on a psoriatic arthritis data set concerning functional disability.

## Keywords

Semicontinuous data, two-part models, overall marginal mean, patient-specific inference, serial correlation, psoriatic arthritis

## 1 Introduction

Semicontinuous data arise when the outcome is a mixture of true zeros and continuously distributed positive values.<sup>1</sup> Some examples in the literature have included average daily alcohol consumption,<sup>1</sup> hospital lengths of stay<sup>2</sup> and medical expenditures.<sup>3,4</sup> In these situations, and more generally, it is natural to view the outcome as a result of two processes, the first determines if the outcome is zero, and if not the second determines the positive value. Two-part models are therefore convenient for the analysis of semicontinuous data and have been used extensively. Recently, Smith et al.<sup>3,4</sup> considered the interesting notion of reparameterising the mean of the positive values in terms of the overall mean, which is arguably a more justified target of inference (see Tom et al.<sup>5</sup> and the references therein). We also consider this notion with respect to the overall marginal mean in our framework.

Two-part marginal models and two-part mixed models have both been proposed for the analysis of longitudinal semicontinuous data. The first is motivated by obtaining population-based inference and have been constructed using generalized estimating equations.<sup>6</sup> The second is more convenient when patient-specific inference

MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

### Corresponding author:

Sean Yiu, School of Clinical Medicine, University of Cambridge, Robinson Way, Cambridge CB2 0SR, UK.

Email: sean.yiu@mrc-bsu.cam.ac.uk

is of interest and are constructed by incorporating correlated patient-specific random effects in both parts of the model.<sup>7</sup> This paper focuses on the two-part mixed modelling approach, although considerations are provided on how population-based inference can be obtained.

In some situations, correlated patient-specific random effects models will not provide an adequate fit to the data. This may especially be the case when the lengths of follow-up are relatively long. Here, it may be less plausible to assume that patients can only have consistently high or low outcomes throughout their entire follow-up. In terms of the correlation structure, it may not be reasonable to assume constant correlation between outcomes from the same patient regardless of their gap times (which is induced by patient-specific random effects). Flexible two-part models that allow for random changes in the trajectory through serially correlated stochastic processes may then be more plausible and these have been proposed in the literature. Albert and Shen<sup>8</sup> and Ghosh and Albert<sup>9</sup> proposed two-part mixed models that consisted of correlated Gaussian processes and random walks (in addition to correlated patient-specific random effects), respectively, in both parts of the model. Albert and Shen<sup>8</sup> demonstrated, through their application and a simulation study, that overall conditional means may suffer from bias if serial correlation (which is not captured by patient-specific random effects) is present but ignored. It is also worth noting, both models incorporating stochastic processes provided considerable improvements of fit to their data.

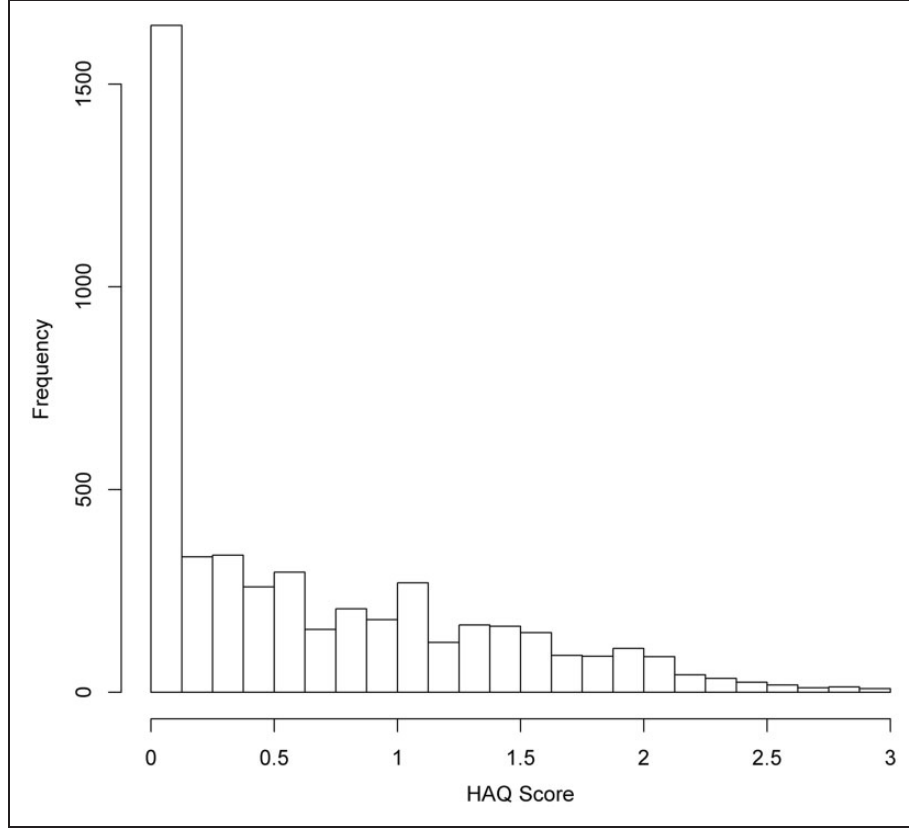
A main drawback of fitting models with stochastic processes is the computationally intensive nature of the model fitting procedure. The primary difficulty results from the following feature: if a patient has  $m_i$  observations, then a model consisting of correlated stochastic processes in each part of the model will require  $2m_i$  integrations to evaluate the marginal likelihood contribution from that patient (assuming, as is usual, the stochastic processes are realised at the observation times). For manageable values of  $m_i$ , Albert and Shen<sup>8</sup> and Ghosh and Albert<sup>9</sup> have developed methods based on a Monte Carlo Expectation Maximization algorithm and Markov chain Monte Carlo, respectively, to evaluate the marginal likelihood. Both of these procedures can be computationally intensive, with the former also requiring standard errors of parameter estimates to be computed by bootstrap. The primary aim of this paper is to demonstrate, using a property of the multivariate normal distribution and the standard marginal cumulative distribution function identity, how a marginal likelihood can be obtained in terms of the cumulative distribution function of a multivariate normal distribution. Implicitly, because it is possible to efficiently evaluate the cumulative distribution function of a multivariate normal distribution, maximum likelihood estimation can be used to obtain parameter estimates and (asymptotic) standard errors (from the observed information matrix) of model parameters.

The rest of this paper is organised as follows. In Section 2, the motivating application concerning functional disability in psoriatic arthritis is introduced. Section 3 describes the flexible two-part modelling framework of Albert and Shen<sup>8</sup> and Ghosh and Albert<sup>9</sup> (including additional comments regarding implementation). Section 4 proposes an efficient maximum likelihood estimation procedure for the models in Section 3. Section 5 applies the methodology in Section 4 to the data described in Section 2. While retaining the flexibility of using stochastic processes models and the practicality of the proposed efficient implementation procedure, Section 6 extends the modelling framework of Section 3 to allow for the direct modelling of the overall marginal mean. Finally, concluding remarks are made in Section 7.

## 2 Functional disability in psoriatic arthritis

Psoriatic arthritis (PsA) is an inflammatory arthritis associated with the skin condition psoriasis. Because of both skin and joint involvement of the disease, PsA can result in patients having severe physical functional disability. The dominant measure of functional disability in PsA, as well as in many other disease areas,<sup>10</sup> is the self-reported health assessment questionnaire (HAQ). This produces an essentially continuous measure<sup>11–14</sup> between zero, representing no disability, and three, representing severe disability.

The HAQ scores of 698 patients observed longitudinally at the University of Toronto PsA clinic were considered for this analysis. Figure 1 shows the frequencies of HAQ scores from these patients. From Figure 1, it is evident that a large proportion of zeros exist in this data set ( $1526/4811 = 0.32$ ). The clumping at zero, together with the continuous distributed outcomes for the non-zero values, suggests that the HAQ score can be viewed as a semicontinuous outcome. Su et al.<sup>12,13</sup> considered two-part models with patient-specific random effects for analysing an earlier version of this PsA data set. In this paper, we relax the assumption of constant patient-specific random effects to patient-specific stochastic processes and consider the extent to which they improve understanding of the disability process. This includes making easy interpretable inference on the overall marginal mean HAQ scores, a concept that has not been considered before with stochastic processes models



**Figure 1.** Frequencies of HAQ scores in our data.

(see Section 6 for more details). On average, patients had 6.89 clinic visits (ranging from 2 to 20) with mean inter-visit and follow-up times of 1 year and 5 months (standard deviation (SD) of 1 year and 1 month) and 8 years and 3 months (SD of 5 years and 10 months), respectively.

### 3 Model

Let  $Y_{ij}$  ( $i = 1, \dots, N$ ) denote the semicontinuous response from patient  $i$  at time  $t_{ij}$  ( $j = 1, \dots, m_i$ ), where  $t_{ij}$  represents the time of the  $j$ th observation from patient  $i$ . Because of true zeros, it is natural to decompose the response into

$$U_{ij} = \begin{cases} 1 & : Y_{ij} > 0 \\ 0 & : Y_{ij} = 0 \end{cases}$$

and  $g(Y_{ij})|Y_{ij} > 0$ , where  $g(\cdot)$  is a monotonic function such that  $g(0) = 0$  and  $g(Y_{ij})|Y_{ij} > 0$  is positive and approximately Gaussian with constant variance  $\sigma^2$ . For convenience, the model for  $U_{ij}$  is referred to as the binary component, while the model for  $g(Y_{ij})|Y_{ij} > 0$  is referred to as the continuous component.

We now describe the flexible modelling framework. Let  $\mathbf{X}_{ij}$  and  $\mathbf{Z}_{ij}$  be column vectors of covariates that influence the probability of  $Y_{ij} > 0$  and the mean of  $g(Y_{ij})|Y_{ij} > 0$ , respectively. Then conditional on correlated patient-specific random effects ( $B_i^r = b_i^r, C_i^r = c_i^r$ ) and correlated stochastic processes  $\{B_i^s(t_{ij}) = b_{ij}^s, C_i^s(t_{ij}) = c_{ij}^s\}$ , where the random effects are assumed independent of the stochastic processes, we model  $U_{ij}$  as Bernoulli with response probability

$$\mathbb{P}(U_{ij} = 1 | b_{ij}^s, b_i^r) = \Phi(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + b_{ij}^s + b_i^r) \quad (1)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of a standard Gaussian distribution (i.e. probit model), and  $[g(Y_{ij})|Y_{ij} > 0; c_{ij}^s, c_i^r]$  as Gaussian with mean  $\mathbf{Z}_{ij}^\top \boldsymbol{\gamma} + c_{ij}^s + c_i^r$  and constant variance  $\sigma^2$  (i.e. linear mixed effect model

on  $g(Y_{ij})|Y_{ij} > 0$ ). Here,  $\beta$  and  $\gamma$  are column vectors of regression coefficients. The patient-specific random effects  $(B_i^r, C_i^r)$  allow patients to have a consistently high or low probability of having disability and a consistently high or low mean for the non-zero HAQ scores across time. While the patient-specific stochastic processes  $\{B_i^s(t_{i1}), \dots, B_i^s(t_{im_i}), C_i^s(t_{i1}), \dots, C_i^s(t_{im_i})\}$  can capture serial correlation and non-predictable changes in unobserved heterogeneity.<sup>9</sup>

We assume  $\{B_i^r, C_i^r\}$  follows a bivariate normal distribution with mean vector zero and

$$\text{Var}(B_i^r) = \sigma_b^2, \quad \text{Var}(C_i^r) = \sigma_c^2, \quad \text{Cov}(B_i^r, C_i^r) = \sigma_b \sigma_c \rho \quad (2)$$

where  $\sigma_b^2$  and  $\sigma_c^2$  are variance parameters and  $\rho$  is the correlation between  $B_i^r$  and  $C_i^r$ . Furthermore, we consider two classes of stochastic processes for  $\{B_i^s(t), C_i^s(t)\}$  that are subsequently described. For convenience, define  $B_i(t) = B_i^r + B_i^s(t)$  and  $C_i(t) = C_i^r + C_i^s(t)$ , i.e. the patient-specific random effects  $B_i^r$  and  $C_i^r$  are absorbed into the stochastic processes  $B_i(t)$  and  $C_i(t)$ , respectively, and let the covariance matrix of  $\{B_i(t_{i1}), \dots, B_i(t_{im_i}), C_i(t_{i1}), \dots, C_i(t_{im_i})\}$  be

$$\Sigma_i = \begin{pmatrix} \Sigma_{ib} & \Sigma_{ibc} \\ \Sigma_{ibc} & \Sigma_{ic} \end{pmatrix}.$$

### 3.1 Correlated Gaussian processes

The first and most general model that we consider is defined when  $\{B_i^s(t), C_i^s(t)\}$  are correlated stationary Gaussian processes. That is the model proposed by Albert and Shen<sup>8</sup>

$$\begin{aligned} \text{Cov}\{B_i^s(t_{ij}), B_i^s(t_{ik})\} &= \sigma_{gb}^2 \rho_{gb}^{|t_{ij}-t_{ik}|}, & \text{Cov}\{C_i^s(t_{ij}), C_i^s(t_{ik})\} &= \sigma_{gc}^2 \rho_{gc}^{|t_{ij}-t_{ik}|} \\ \text{Cov}\{B_i^s(t_{ij}), C_i^s(t_{ik})\} &= \sigma_{gb} \sigma_{gc} \rho_g \rho_{gbc}^{|t_{ij}-t_{ik}|} \end{aligned} \quad (3)$$

where  $\sigma_{gb}^2$  and  $\sigma_{gc}^2$  are variance parameters,  $\rho_g$  is the correlation between the Gaussian processes at the same time point, and  $\rho_{gb}$ ,  $\rho_{gc}$ ,  $\rho_{gbc}$  are the degradation parameters governing the serial correlation within and between processes, respectively. Following Albert and Shen,<sup>8</sup> the processes  $B_i(t)$  and  $C_i(t)$  are taken to be exchangeable Ornstein-Uhlenbeck (EOU) processes, and the model containing these processes is called the general model, i.e. (1–3). Some special cases of the general model are

- Shared EOU process model when  $c_{ij} = \theta b_{ij}$ ,
- Correlated OU processes model when  $\sigma_b^2 \equiv \sigma_c^2 \equiv \rho \equiv 0$ ,
- Shared OU process model when  $c_{ij}^s = \theta b_{ij}^s$  and  $\sigma_b^2 \equiv \sigma_c^2 \equiv \rho \equiv 0$ ,
- Correlated random effects model when  $\sigma_{gb}^2 \equiv \sigma_{gc}^2 \equiv \rho_g \equiv \rho_{gb} \equiv \rho_{gc} \equiv \rho_{gbc} \equiv 0$ ,
- Shared random effect model when  $c_i^r = \theta b_i^r$  and  $\sigma_b^2 \equiv \sigma_c^2 \equiv \rho_g \equiv \rho_{gb} \equiv \rho_{gc} \equiv \rho_{gbc} \equiv 0$ ,

where  $\theta$  is a parameter to be estimated.

#### 3.1.1 Remarks on $\rho_{gb}$ , $\rho_{gc}$ and $\rho_{gbc}$

Although the general model is very flexible, it will not always be mathematically valid. Let the covariance matrices  $\Sigma_{ib}^s$ ,  $\Sigma_{ic}^s$  and  $\Sigma_{ibc}^s$  have  $(j, k)$ th entry  $\sigma_{gb}^2 \rho_{gb}^{|t_{ij}-t_{ik}|}$ ,  $\sigma_{gc}^2 \rho_{gc}^{|t_{ij}-t_{ik}|}$  and  $\sigma_{gb} \sigma_{gc} \rho_g \rho_{gbc}^{|t_{ij}-t_{ik}|}$  respectively, i.e. described by equation (3). If  $\rho_{gb}$ ,  $\rho_{gc}$  and  $\rho_{gbc}$  are unconstrained (as specified by Albert and Shen<sup>8</sup>), the matrix  $\Sigma_i^s$  where

$$\Sigma_i^s = \begin{pmatrix} \Sigma_{ib}^s & \Sigma_{ibc}^s \\ \Sigma_{ibc}^s & \Sigma_{ic}^s \end{pmatrix}$$

will not in general be a valid covariance matrix since  $\Sigma_i^s$ , although symmetric, is not constrained to be positive semi-definite and therefore  $B_i^s(t)$  and  $C_i^s(t)$  will not necessarily form a jointly Gaussian process. The primary difficulty results when  $\rho_g$  (the correlation between  $B_i^s(t)$  and  $C_i^s(t)$  at each time  $t$ ) is close to one because the processes  $B_i^s(t)$  and  $C_i^s(t)$  are similar and therefore it will not be plausible for them to degrade at vastly different rates (i.e. for  $\rho_{gb}$ ,  $\rho_{gc}$  and  $\rho_{gbc}$  to be vastly different). A reasonable approximation in this situation would be to constrain the degradation and cross degradation parameters to be same, specifically  $\rho_{gb} \equiv \rho_{gc} \equiv \rho_{gbc} \equiv \rho_{g1}$ . This constraint would then enforce  $\Sigma_i^s$  to be a valid covariance matrix since the Schur

component  $\Sigma_{ib}^s - \Sigma_{ibc}^s(\Sigma_{ic}^s)^{-1}\Sigma_{ibc}^s \equiv \sigma_{gb}^2(1 - \rho_g^2)\Sigma_i^s(\rho_{g1})$ , where  $\Sigma_i^s(\rho_{g1})$  has  $(j, k)$ th entry  $\rho_{g1}^{|t_{ij}-t_{ik}|}$ , is constrained to be positive semi-definite. The resulting correlation structure would then be

$$\begin{aligned}\text{Cor}\{B_i^s(t_{ij}), B_i^s(t_{ik})\} &= \text{Cor}\{C_i^s(t_{ij}), C_i^s(t_{ik})\} = \rho_{g1}^{|t_{ij}-t_{ik}|} \\ \text{Cor}\{B_i^s(t_{ij}), C_i^s(t_{ik})\} &= \rho_g \rho_{g1}^{|t_{ij}-t_{ik}|}\end{aligned}\quad (4)$$

In the motivating application,  $\rho_g$  was estimated close to one. Slight deviations from the correlation structure described by equation (4) (for example  $\rho_{gb} \equiv \rho_{gc} \equiv \rho_{g1}$  and  $\rho_{gbc} = \rho_{g1}\delta$  where  $\delta \in (0, 1)$ ) resulted in non-positive semi-definite matrices for various  $\Sigma_i$ , and therefore the model fitting procedure was problematic. Note that a further simplification would be to constrain  $\rho_g = 1$  (in addition to  $\rho_{g1} = \rho_{gb}$ ), this would result in the shared EOU process model. If, however,  $\rho_g$  takes a smaller value, and therefore the two Gaussian processes are less correlated, it would then be more plausible for the Gaussian processes to degrade at different rates. Hence, having unconstrained degradation parameters will likely be less problematic.

For completeness, note that

$$\begin{aligned}\{\Sigma_{ib}\}_{jk} &= \sigma_b^2 + \sigma_{gb}^2 \rho_{gb}^{|t_{ij}-t_{ik}|} \\ \{\Sigma_{ic}\}_{jk} &= \sigma_c^2 + \sigma_{gc}^2 \rho_{gc}^{|t_{ij}-t_{ik}|} \\ \{\Sigma_{ibc}\}_{jk} &= \sigma_b \sigma_c \rho + \sigma_{gb} \sigma_{gc} \rho_g \rho_{gbc}^{|t_{ij}-t_{ik}|}\end{aligned}\quad (5)$$

### 3.2 Correlated random walks

The second model structure that we consider is defined when  $\{B_i^s(t), C_i^s(t)\}$  are correlated continuous-time random walks. That is the model proposed by Ghosh and Albert.<sup>9</sup> Specifically, define sequentially  $\{B_i(t_{ij+1}), C_i(t_{ij+1})\} | \{B_i(t_{ij}) = b_{ij}, C_i(t_{ij}) = c_{ij}\}$  to be bivariate normal with mean  $(b_{ij}, c_{ij})$  and covariance matrix

$$\begin{pmatrix} \sigma_{wb}^2(t_{ij+1} - t_{ij}) & \sigma_{wb}\sigma_{wc}\rho_w(t_{ij+1} - t_{ij}) \\ \sigma_{wb}\sigma_{wc}\rho_w(t_{ij+1} - t_{ij}) & \sigma_{wc}^2(t_{ij+1} - t_{ij}) \end{pmatrix}$$

In addition  $(b_{i1}, c_{i1}) = (b_i^r, c_i^r)$  are initiated at realisations of the patient-specific random effects. Here,  $\sigma_{wb}^2$ ,  $\sigma_{wc}^2$  and  $\rho_w$  are variance and correlation parameters that quantify serial correlation (both within and across processes). This model will be denoted by a correlated random walks (CRW) model and it contains as special cases.

- Shared random walk model when  $c_{ij} = \theta b_{ij}$ ,
- Correlated random effects model when  $\sigma_{wb}^2 \equiv \sigma_{wc}^2 \equiv \rho_w \equiv 0$ ,
- Shared random effect model when  $c_i^r = \theta b_i^r$  and  $\sigma_{wb}^2 \equiv \sigma_{wc}^2 \equiv \rho_w \equiv 0$ .

Although the CRW model is less flexible than the general model, it has the advantage, from its sequential construction, of always being well defined even when the parameters are unconstrained (apart from the usual constraint that correlation parameters have modulus less than or equal to unity). Moreover

$$\begin{aligned}\{\Sigma_{ib}\}_{jk} &= \sigma_b^2 + \sigma_{wb}^2 \{\min(t_{ij}, t_{ik}) - t_{i1}\} \\ \{\Sigma_{ic}\}_{jk} &= \sigma_c^2 + \sigma_{wc}^2 \{\min(t_{ij}, t_{ik}) - t_{i1}\} \\ \{\Sigma_{ibc}\}_{jk} &= \sigma_b \sigma_c \rho + \sigma_{wb} \sigma_{wc} \rho_w \{\min(t_{ij}, t_{ik}) - t_{i1}\}\end{aligned}\quad (6)$$

## 4 Efficient maximum likelihood estimation procedure for stochastic processes models

This section describes our efficient maximum likelihood estimation procedure for the flexible models described in Section 3. Firstly, in Section 4.1, we describe a generic likelihood function for all of the described models. The multivariate normal identity that can be used to evaluate certain multi-dimensional integrals in terms of a multivariate normal cumulative distribution function is introduced in Section 4.2. Finally, in Section 4.3, we outline how to apply the multivariate normal identity in Section 4.2 to the generic likelihood function in

Section 4.1, thus culminating in a computationally efficient likelihood. For completeness, we also provide computational simplifications for correlated stochastic processes models in the appendix.

#### 4.1 Likelihoods

For ease of exposition, we describe the likelihood contribution from patient  $i$ . The likelihood can then be obtained by taking the product of all likelihood contributions from each patient. Firstly, we consider models that contain two (correlated) stochastic processes. For these models, the likelihood contribution from patient  $i$  is

$$L_i(\Theta) = \int_{b_i} \int_{c_i} \left[ \prod_{j=1}^{m_i} \Phi(X_{ij}^\top \beta + b_{ij})^{u_{ij}} \{1 - \Phi(X_{ij}^\top \beta + b_{ij})\}^{1-u_{ij}} \right] \times \left[ \prod_{j=1}^{m_i} \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(g(y_{ij}) - \mathbf{Z}_{ij}^\top \gamma - c_{ij})^2}{2\sigma^2} \right\} \right]^{u_{ij}} \right] \phi^{(2m_i)}(\mathbf{b}_i, \mathbf{c}_i; \mathbf{0}, \Sigma_i) d\mathbf{b}_i d\mathbf{c}_i \quad (7)$$

where  $\Theta$  is a vector comprising all of the unknown parameters,  $\mathbf{b}_i = (b_{i1}, \dots, b_{im_i})^\top$ ,  $\mathbf{c}_i = (c_{i1}, \dots, c_{im_i})^\top$ ,  $\phi^{(m)}(\cdot; \mu, \Sigma)$  is an  $m$  dimensional multivariate normal density with mean vector  $\mu$  and covariance matrix  $\Sigma$ , and  $\Sigma_i$  is defined by either equation (5) or equation (6). Similarly, for models containing a single stochastic process (i.e. shared process models), the likelihood contribution from patient  $i$  is

$$L_i(\Theta) = \int_{b_i} \left[ \prod_{j=1}^{m_i} \Phi(X_{ij}^\top \beta + b_{ij})^{u_{ij}} \{1 - \Phi(X_{ij}^\top \beta + b_{ij})\}^{1-u_{ij}} \right] \times \left[ \prod_{j=1}^{m_i} \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(g(y_{ij}) - \mathbf{Z}_{ij}^\top \gamma - \theta b_{ij})^2}{2\sigma^2} \right\} \right]^{u_{ij}} \right] \phi^{(m_i)}(\mathbf{b}_i; \mathbf{0}, \Sigma_{ib}) d\mathbf{b}_i \quad (8)$$

where  $\Sigma_{ib}$  can again be obtained from equation (5) or equation (6). We now define our generic likelihood contribution from patient  $i$  which encompasses all of the described models. Throughout we apply the following notation:  $\mathbf{0}$  and  $\mathbf{1}$  are  $m_i \times 1$  vectors with all entries being zero and one respectively,  $\text{diag}(\mathbf{v})$  is a matrix with diagonal elements  $\mathbf{v}$  and zero otherwise, and  $\mathbf{I}_d$  is a  $d \times d$  identity matrix. We also follow the convention that binary operations with a scalar and vector or matrix argument and unary operations with a vector argument are performed element-wise. In matrix form, we have

$$L_i(\Theta) = \int_{l_i} \Phi^{(m_i)}(A_{i1}\mu_{ib} + A_{i2}l_i; \mathbf{0}, \mathbf{I}_{m_i}) \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\sum_{j=1}^{m_i} u_{ij}} \times \exp \left\{ -\frac{(g(\mathbf{y}_i) - A_{i3}\mu_{ic} - A_{i4}l_i)^\top (g(\mathbf{y}_i) - A_{i3}\mu_{ic} - A_{i4}l_i)}{2\sigma^2} \right\} \phi^{\{\dim(l_i)\}}(l_i; \mathbf{0}, \Sigma_{il}) dl_i \quad (9)$$

where  $\mathbf{y}_i = (y_{i1}, \dots, y_{im_i})^\top$ ,  $\mu_{ib} = X_i\beta$ ,  $\mu_{ic} = \mathbf{Z}_i\gamma$ ,  $X_i = (X_{i1}, \dots, X_{im_i})^\top$ ,  $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{im_i})^\top$ , and  $A_{i1} = \text{diag}(2\mathbf{u}_i - \mathbf{1})$ ,  $A_{i3} = \text{diag}(\mathbf{u}_i)$  are  $m_i \times m_i$  matrices with  $\mathbf{u}_i = (u_{i1}, \dots, u_{im_i})^\top$ . Here  $\Phi^{(d)}(\cdot; \mathbf{0}, \Sigma)$  represents the distribution function of  $\phi^{(d)}(\cdot; \mathbf{0}, \Sigma)$  and  $l_i$  is a (to be specified) column vector of random effects. Note that equation (9) has resulted from repeated application of the identity  $1 - \Phi(x) = \Phi(-x)$ .

The likelihood contribution from patient  $i$ ,  $L_i(\Theta)$ , is then obtained by specifying the vector of random effects  $l_i$  and its covariance matrix  $\Sigma_{il}$  together with the  $m_i \times \dim(l_i)$  matrices  $A_{i2}$  and  $A_{i4}$  which describe how the random effects act on the binary and continuous components of the model. For (7),  $l_i = (\mathbf{b}_i, \mathbf{c}_i)$ ,  $A_{i2} = (\text{diag}(2\mathbf{u}_i - \mathbf{1}), \text{diag}(\mathbf{0}))$  and  $A_{i4} = (\text{diag}(\mathbf{0}), \text{diag}(\mathbf{u}_i))$ . While for equation (8),  $l_i = \mathbf{b}_i$ ,  $A_{i2} = \text{diag}(2\mathbf{u}_i - \mathbf{1})$  and  $A_{i4} = \text{diag}(\theta\mathbf{u}_i)$ . Similarly, for the correlated random effects model,  $l_i = (b_i, c_i)$ ,  $A_{i2} = (2\mathbf{u}_i - \mathbf{1}, \mathbf{0})$  and  $A_{i4} = (\mathbf{0}, \mathbf{u}_i)$ , and for the shared random effect model,  $l_i = b_i$ ,  $A_{i2} = 2\mathbf{u}_i - \mathbf{1}$  and  $A_{i4} = \theta\mathbf{u}_i$ .

#### 4.2 Multivariate normal identity

In order to evaluate the likelihood described by equation (9), we derive a multivariate normal identity that makes use of a property of the multivariate normal distribution and the standard marginal cumulative distribution function identity. Firstly, suppose that  $\omega = (\omega_1, \omega_2)^\top$  follows a multivariate normal distribution with mean



vector  $(\mathbf{0}, \boldsymbol{\eta})^\top$  where  $\boldsymbol{\omega}_1$  and  $\mathbf{0}$  are  $k_1 \times 1$  vectors and  $\boldsymbol{\omega}_2$  and  $\boldsymbol{\eta}$  are  $k_2 \times 1$  vectors, respectively. Furthermore, suppose that the covariance matrix of  $\boldsymbol{\omega}$  is the  $(k_1 + k_2) \times (k_1 + k_2)$  matrix  $\boldsymbol{\Sigma}$  where the first  $k_1$  rows of  $\boldsymbol{\Sigma}$  is the  $k_1 \times (k_1 + k_2)$  matrix  $(\boldsymbol{\Sigma}_{22}, \boldsymbol{\Sigma}_{12}^\top)$  and the remaining  $k_2$  rows of  $\boldsymbol{\Sigma}$  is the  $k_2 \times (k_1 + k_2)$  matrix  $(\boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{11})$  respectively. It is a well-known result that  $\phi^{(k_1+k_2)}(\boldsymbol{\omega}; (\mathbf{0}, \boldsymbol{\eta})^\top, \boldsymbol{\Sigma}) = \phi^{(k_1)}(\boldsymbol{\omega}_1; \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{\omega}_2 - \boldsymbol{\eta}), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}) \phi^{(k_2)}(\boldsymbol{\omega}_2; \boldsymbol{\eta}, \boldsymbol{\Sigma}_{11})$  where the right-hand side is the product of the conditional density of  $\boldsymbol{\omega}_1 | \boldsymbol{\omega}_2$  and the marginal density of  $\boldsymbol{\omega}_2$ . By applying the standard marginal cumulative distribution function identity  $F_{\boldsymbol{\omega}_1}(\boldsymbol{\omega}_1) = \int_{\boldsymbol{\omega}_2} F_{\boldsymbol{\omega}_1 | \boldsymbol{\omega}_2}(\boldsymbol{\omega}_1 | \boldsymbol{\omega}_2) f_{\boldsymbol{\omega}_2}(\boldsymbol{\omega}_2) d\boldsymbol{\omega}_2$  where the integrand is based on the right-hand side of the above result, we obtain the multivariate normal identity:

$$\Phi^{(k_1)}(\boldsymbol{\omega}_1; \mathbf{0}, \boldsymbol{\Sigma}_{22}) = \int_{\boldsymbol{\omega}_2} \Phi^{(k_1)}(\boldsymbol{\omega}_1 - \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{\omega}_2 - \boldsymbol{\eta}); \mathbf{0}, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}) \phi^{(k_2)}(\boldsymbol{\omega}_2; \boldsymbol{\eta}, \boldsymbol{\Sigma}_{11}) d\boldsymbol{\omega}_2 \quad (10)$$

by noting that the marginal distribution of  $\boldsymbol{\omega}_1$  is multivariate normal with mean vector  $\mathbf{0}$  and covariance matrix  $\boldsymbol{\Sigma}_{22}$ .

Returning to the application, the general idea is to rearrange equation (9) to take the form of the right-hand side of equation (10), and then to use equation (10) to compute the integrations over the random effects in terms of an  $m_i$  dimensional normal cumulative distribution function. Because there exists efficient implementations of the multivariate normal cumulative distribution function, this approach will allow for the efficient computation of the generic likelihood. We note that Barrett et al.<sup>15</sup> used equation (10) to obtain computationally efficient likelihoods of flexible models that jointly consider longitudinal and time to event outcomes. Equation (10) also arises frequently in results concerning the multivariate skew normal distribution.<sup>16–19</sup>

### 4.3 Re-expressing the likelihoods

This section demonstrates how equation (9) (the likelihood contribution from patient  $i$ ) can be re-expressed. We firstly consider the integrand terms resulting from the continuous component and random effects. That is

$$\left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\sum_{j=1}^{m_i} u_{ij}} \exp \left\{ - \frac{(g(\mathbf{y}_i) - \mathbf{A}_{i3}\boldsymbol{\mu}_{ic} - \mathbf{A}_{i4}\mathbf{l}_i)^\top (g(\mathbf{y}_i) - \mathbf{A}_{i3}\boldsymbol{\mu}_{ic} - \mathbf{A}_{i4}\mathbf{l}_i)}{2\sigma^2} \right\} \phi^{\{\dim(\mathbf{l}_i)\}}(\mathbf{l}_i; \mathbf{0}, \boldsymbol{\Sigma}_{il}) \quad (11)$$

By completing the square in  $\mathbf{l}_i$  (see the appendix for more details), equation (11) can be rearranged as

$$L_{i1} \phi^{\{\dim(\mathbf{l}_i)\}}(\mathbf{l}_i; \mathbf{h}_i, \mathbf{H}_i^{-1}) \quad (12)$$

where

$$\begin{aligned} \mathbf{H}_i &= \mathbf{A}_{i4}^\top \mathbf{A}_{i4} / \sigma^2 + (\boldsymbol{\Sigma}_{il})^{-1} \\ \mathbf{h}_i &= \mathbf{H}_i^{-1} \mathbf{A}_{i4}^\top (g(\mathbf{y}_i) - \mathbf{A}_{i3}\boldsymbol{\mu}_{ic}) / \sigma^2 \end{aligned} \quad (13)$$

and

$$L_{i1} = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\sum_{j=1}^{m_i} u_{ij}} \frac{1}{|\boldsymbol{\Sigma}_{il} \mathbf{H}_i|^{1/2}} \exp \left\{ - \frac{(g(\mathbf{y}_i) - \mathbf{A}_{i3}\boldsymbol{\mu}_{ic})^\top (g(\mathbf{y}_i) - \mathbf{A}_{i3}\boldsymbol{\mu}_{ic})}{2\sigma^2} + \frac{\mathbf{h}_i^\top \mathbf{H}_i \mathbf{h}_i}{2} \right\} \quad (14)$$

is independent of  $\mathbf{l}_i$ . Substituting equation (12) into equation (9), we consider the integral (ignoring  $L_{i1}$ )

$$\int_{\mathbf{l}_i} \Phi^{(m_i)}(\mathbf{A}_{i1}\boldsymbol{\mu}_{ib} + \mathbf{A}_{i2}\mathbf{l}_i; \mathbf{0}, \mathbf{I}_{m_i}) \phi^{\{\dim(\mathbf{l}_i)\}}(\mathbf{l}_i; \mathbf{h}_i, \mathbf{H}_i^{-1}) d\mathbf{l}_i. \quad (15)$$

We can re-express the argument and covariance matrix of the multivariate normal distribution function in equation (15) as

$$\begin{aligned} \mathbf{A}_{i1}\boldsymbol{\mu}_{ib} + \mathbf{A}_{i2}\mathbf{l}_i &= \mathbf{A}_{i1}\boldsymbol{\mu}_{ib} + \mathbf{A}_{i2}\mathbf{h}_i - (-\mathbf{H}_i^{-1} \mathbf{A}_{i2}^\top)^\top \mathbf{H}_i(\mathbf{l}_i - \mathbf{h}_i) \\ \mathbf{I}_{m_i} &= \mathbf{I}_{m_i} + \mathbf{A}_{i2}\mathbf{H}_i^{-1} \mathbf{A}_{i2}^\top - (-\mathbf{H}_i^{-1} \mathbf{A}_{i2}^\top)^\top \mathbf{H}_i(-\mathbf{H}_i^{-1} \mathbf{A}_{i2}^\top) \end{aligned} \quad (16)$$

Therefore equation (15), after applying the multivariate normal identity (described equation (10)), is equivalent to

$$\Phi^{(m_i)}(A_{i1}\mu_{ib} + A_{i2}h_i; \mathbf{0}, I_{m_i} + A_{i2}H_i^{-1}A_{i2}^\top) \quad (17)$$

Based on the above expressions, the likelihood contribution from patient  $i$  can now be re-expressed as

$$\begin{aligned} L_i(\Theta) = & \Phi^{(m_i)}(A_{i1}\mu_{ib} + A_{i2}h_i; \mathbf{0}, I_{m_i} + A_{i2}H_i^{-1}A_{i2}^\top) \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\sum_{j=1}^{m_i} u_{ij}} \\ & \times \frac{1}{|\Sigma_{il}H_i|^{1/2}} \exp \left\{ -\frac{(g(y_i) - A_{i3}\mu_{ic})^\top (g(y_i) - A_{i3}\mu_{ic})}{2\sigma^2} + \frac{h_i^\top H_i h_i}{2} \right\} \end{aligned} \quad (18)$$

where

$$\begin{aligned} A_{i1} &= \text{diag}(2\mathbf{u}_i - \mathbf{1}) \\ A_{i3} &= \text{diag}(\mathbf{u}_i) \\ H_i &= A_{i4}^\top A_{i4} / \sigma^2 + (\Sigma_{il})^{-1} \\ h_i &= H_i^{-1} A_{i4}^\top (g(y_i) - A_{i3}\mu_{ic}) / \sigma^2 \end{aligned} \quad (19)$$

and  $\Sigma_{il} = \Sigma_i$  or  $\Sigma_{ib}$  with  $A_{i2}$  and  $A_{i4}$  defined by the specified model.

From equations (18) and (19), it is now evident that evaluating the integrations involved in  $L_i(\Theta)$  reduces to computing the cumulative distribution function of a multivariate normal distribution. This can be performed efficiently, for example by using the R<sup>20</sup> package `mnormt`.<sup>21</sup> The model fitting procedure is then completed by maximizing the log-likelihood, for example by using the BFGS<sup>22</sup> optimization technique, to obtain parameter estimates and asymptotic standard errors (from the observed Fisher information matrix) of model parameters.

## 5 Application: Patient-specific inference

Using the estimation procedure described in Section 4, we demonstrate how patient-specific inference on the probability of being disabled and the transformed mean HAQ score conditional on disability can be obtained. Specifically, how a unit change in covariate values impacts these quantities for any specific patient. We consider the covariate effects of the number of clinically damaged joints (time-dependent), the number of actively inflamed joints (time-dependent), sex (coded as 1 for males and 0 for females), arthritis duration in years (time-dependent), and age at onset of arthritis in years (standardise). Following Su et al.,<sup>12,13</sup> no transformation was applied to the non-zero HAQ scores, i.e.  $g(y) = y$ .

Initially, models with two stochastic processes were fitted to the HAQ data. This resulted in large estimated correlation parameters between the random effects (i.e.  $\rho \approx 1$ ) and stochastic processes for both the correlated Gaussian processes and random walks cases (i.e.  $\rho_g$  and  $\rho_w \approx 1$ ). These results therefore suggested a single stochastic process would be sufficient for describing the data. The shared EOU model was then fitted. However, the analysis provided evidence for model over-parameterisation as  $\hat{\sigma}^2$  appeared to converge at virtually zero and a positive-definite observed Fisher information matrix could not be attained (even when a considerably smaller tolerance level than the default was specified for the computation of multivariate normal probabilities). We therefore considered the shared random walk and OU process models, and for comparative purposes, the shared random effect model. The models containing stochastic processes were fitted using the likelihood described by equations (18) and (19), while the shared random effect model was fitted using numerical integration (since only a single integration per patient is required). The same parameter estimates for the shared random effect model were obtained when equations (18) and (19) were used in the model fitting procedure.

Table 1 presents the results of the fitted models. Across the models, the covariate effects on the mean conditional on disability are seen to be relatively similar as the confidence intervals generally overlap. In addition, the models are in agreement with regard to the association of each covariate apart from arthritis duration. Arthritis duration is statistically significant in the shared random effect model but is not statistically significant in the models that incorporate stochastic processes. It is interesting to note that there are strong



**Table 1.** Table displaying patient-specific effects and corresponding 95% Wald intervals on the probability of being disabled and the mean HAQ score conditional on disability.

|                                     | Shared random walk       | Shared OU process       | Shared random effect     |
|-------------------------------------|--------------------------|-------------------------|--------------------------|
| <b>Binary component</b>             |                          |                         |                          |
| Damaged joints                      | 0.031 (0.013, 0.049)     | 0.04 (0.021, 0.059)     | 0.012 (0.002, 0.022)     |
| Active joints                       | 0.16 (0.13, 0.18)        | 0.17 (0.15, 0.2)        | 0.15 (0.13, 0.16)        |
| Sex                                 | −1.72 (−1.9, −1.54)      | −2.17 (−2.92, −1.42)    | −1.34 (−1.65, −1.02)     |
| Arthritis duration                  | 0.042 (0.025, 0.058)     | 0.044 (0.022, 0.065)    | 0.034 (0.023, 0.044)     |
| Age at arthritis onset <sup>a</sup> | 0.56 (0.45, 0.66)        | 0.68 (0.48, 0.88)       | 0.45 (0.3, 0.6)          |
| Intercept                           | 1.77 (1.62, 1.92)        | 1.97 (1.07, 2.87)       | 1.11 (0.81, 1.41)        |
| <b>Continuous component</b>         |                          |                         |                          |
| Damaged joints                      | 0.0065 (0.0032, 0.0097)  | 0.0078 (0.0046, 0.011)  | 0.0033 (0.00086, 0.0058) |
| Active joints                       | 0.021 (0.019, 0.023)     | 0.021 (0.019, 0.023)    | 0.02 (0.018, 0.022)      |
| Sex                                 | −0.29 (−0.34, −0.24)     | −0.35 (−0.46, −0.23)    | −0.29 (−0.37, −0.21)     |
| Arthritis duration                  | 0.0025 (−0.0008, 0.0058) | 0.0035 (−0.001, 0.0081) | 0.0067 (0.0041, 0.0093)  |
| Age at arthritis onset <sup>a</sup> | 0.076 (0.048, 0.1)       | 0.093 (0.046, 0.14)     | 0.086 (0.048, 0.12)      |
| Intercept                           | 0.62 (0.57, 0.66)        | 0.59 (0.44, 0.74)       | 0.63 (0.56, 0.7)         |
| $\theta$                            | 0.2 (0.19, 0.22)         | 0.19 (0.17, 0.21)       | 0.27 (0.24, 0.3)         |
| $\sigma^2$                          | 0.074 (0.069, 0.079)     | 0.066 (0.06, 0.072)     | 0.12 (0.11, 0.12)        |
| $\sigma_b^2$                        | 6.3 (5.64, 7.04)         |                         | 3.29 (2.64, 4.1)         |
| $\sigma_{gb}^2$                     |                          | 10.11 (8.03, 12.74)     |                          |
| $\sigma_{wb}^2$                     | 0.58 (0.52, 0.65)        |                         |                          |
| $\rho_{gb}$                         |                          | 0.95 (0.94, 0.96)       |                          |
| Log-likelihood                      | −3279.08                 | −3282.11                | −3500.48                 |

<sup>a</sup>Denotes the standardised version of the covariate.

agreements regarding the covariate effect of the number of active joints (similar parameter estimates across models and relatively narrow confidence intervals). The models indicate an additional actively inflamed joint will increase the mean HAQ score conditional on disability by approximately 0.21 for any specific patient. For the binary component, the covariate effects are again seen to be relatively similar due to the overlapping confidence intervals. Their interpretation through the direction of association and statistical significance are also consistent across models. The covariate effects from the shared random effect model does, however, consistently demonstrate attenuation to the null when compared to the other models with stochastic processes.

A generalized likelihood ratio test of  $\sigma_{wb}^2 = 0$  and  $\rho_{gb} = 1$  produced  $p$  values of  $< 0.001$  therefore suggesting preference towards the shared random walk and OU process models respectively when compared to the shared random effect model. Since the shared random walk and OU process models contain the same number of parameters, information criteria, such as AIC, would indicate (weakly) that the shared random walk model is preferable. It is also worth noting that the heterogeneity parameter in the binary component (i.e.  $\sigma_b^2$  or  $\sigma_{gb}^2$ ) is significantly lower in the shared random effect model. For this model, this parameter governs both the heterogeneity and correlation due to repeated measurements and therefore in light of greater unaccounted heterogeneity (compared to the models with stochastic processes), less correlation is expected.<sup>23</sup> In the continuous component, where  $\sigma^2$  also accounts for heterogeneity, a smaller difference between the heterogeneity parameters (i.e.  $\theta^2 \sigma_b^2$  or  $\theta^2 \sigma_{gb}^2$ ) is seen; in the order of the models displayed in the table (from right to left), the heterogeneity parameters are 0.24, 0.36 and 0.25, respectively.

## 6 Modelling the overall marginal mean

In many cases, it is of interest to obtain population-based inference in addition/as opposed to patient-specific inference. For example, for strategic public health policy purposes, it would be more clinically meaningful to obtain covariate effects on quantities of interest after averaging over all patients. Currently, the proposed models are parametrised to allow easily interpretable patient-specific covariate effects, those with  $B_i(t_{ij}) = b_{ij}$  and  $C_i(t_{ij}) = c_{ij}$ , to act on the patient-specific mean of the transformed positive values (i.e.  $\mathbb{E}[g(Y_{ij})|Y_{ij} > 0, C_i(t_{ij}) = c_{ij}]$ ) and the patient-specific probability of a having a positive value (i.e.

$\mathbb{P}(U_{ij} = 1 | B_i(t_{ij}) = b_{ij})$ ). However, under this parametrisation, it no longer becomes straightforward to obtain easily interpretable population-level covariate effects on the marginal mean of the transformed positive values (the mean of the transformed positive values after averaging over all  $b_{ij}$  and  $c_{ij}$ , i.e.  $\mathbb{E}[g(Y_{ij}) | Y_{ij} > 0]$ ) since it is a highly non-linear function of the linear predictors in the binary and continuous components.<sup>5</sup> Thus, the effect of a single covariate is generally interpreted by fixing other covariates at certain values.<sup>8</sup> This problem remains even when population-level covariate effects on the overall marginal mean of the transformed values (i.e.  $\mathbb{E}[g(Y_{ij})]$ ) are of primary interest, which has strongly been argued as an important target of inference;<sup>24</sup> it is estimated using data from the same patients over time (unlike  $\mathbb{E}[g(Y_{ij}) | Y_{ij} > 0]$ ) and it is a measure of the undecomposed outcome. We reiterate that in considering the overall marginal mean of the transformed values as a target of inference, we assume that the monotonic transformation function is such that  $g(0) = 0$  and  $g(Y_{ij}) | Y_{ij} > 0$  is positive and approximately Gaussian with constant variance  $\sigma^2$ .

In order to obtain population-based inference on the overall marginal mean of the transformed values, Smith et al.<sup>4</sup> proposed the following model parameterisation

$$\begin{aligned}\mathbb{P}(U_{ij} = 1 | B_i^r = b_i^r) &= g_1(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + b_i^r) \\ \mathbb{E}[g(Y_{ij}) | C_i^r = c_i^r] &= g_2(\mathbf{Z}_{ij}^\top \boldsymbol{\alpha} + c_i^r)\end{aligned}\quad (20)$$

where  $g_1(\cdot)$  and  $g_2(\cdot)$  are monotonic link functions and  $B_i^r, C_i^r$  are, as before, zero mean bivariate normal patient-specific random effects. Recall that transformation and link functions differ in that transformation functions are applied prior to modelling. In their specific context, Smith et al.<sup>4</sup> considered the identity transformation for  $g(\cdot)$  but allowed the positive values of  $Y_{ij}$  to follow a log-skew-normal distribution. Under this parametrisation, for a suitably chosen link function such as  $g_2^{-1}$  being the identity or log link, it is implicit that easily interpretable covariate effects on the overall marginal mean of  $g(Y_{ij})$ ,  $\boldsymbol{\alpha}$ , can now be obtained. Smith et al.<sup>4</sup> implemented this model by using a Bayesian estimation approach with

$$\mathbb{E}[g(Y_{ij}) | Y_{ij} > 0, B_i^r = b_i^r, C_i^r = c_i^r] = \frac{g_2(\mathbf{Z}_{ij}^\top \boldsymbol{\alpha} + c_i^r)}{g_1(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + b_i^r)}$$

specified in the likelihoods defined by equation (7) or equation (8). Note that  $\mathbb{E}[g(Y_{ij}) | Y_{ij} > 0, B_i^r = b_i^r, C_i^r = c_i^r]$  is no longer parametrised to be equivalent to a monotonic function of a linear predictor, as was specified before. While this approach for modelling the overall marginal mean is intuitive, it is clear that the multivariate normal identity in Section 4.2 can no longer be used to compute the integrations over the multi-dimensional random effects in the marginal likelihood. Thus, as mentioned in the introduction, implementation of such models can be computationally challenging, especially for our situation where it would be of interest to consider  $b_{ij}$  and  $c_{ij}$  (i.e. realisations of stochastic processes) instead of  $b_i^r$  and  $c_i^r$  (i.e. realisations of patient-specific random effects) in equation (20).

We now propose another method which would allow easily interpretable covariate effects to act on the overall marginal mean of  $g(Y_{ij})$ . In contrast, this method facilitates the inclusion of stochastic processes because it retains the proposed efficient implementation procedure described in Section 4. To the best of our knowledge, there are no other methods in the literature that facilitates the practical implementation of stochastic processes models for directly modelling the overall marginal mean.

We first begin by computing the overall marginal mean of  $g(Y_{ij})$  when

$$\begin{aligned}\mathbb{P}(U_{ij} = 1 | B_i(t_{ij}) = b_{ij}) &= g_1(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + b_{ij}) \\ \mathbb{E}[g(Y_{ij}) | Y_{ij} > 0, C_i(t_{ij}) = c_{ij}] &= \Delta_{ij} + c_{ij}\end{aligned}\quad (21)$$

where  $\Delta_{ij}$  is a function of covariates (at the  $j$ th visit from patient  $i$ ) and regression coefficients only. That is  $c_{ij}$  is now assumed to act additively on the mean of the transformed positive values and not on the overall mean of the transformed values as is the case in equation (20). For models with two processes, the overall marginal mean of  $g(Y_{ij})$  is defined by

$$\mathbb{E}[g(Y_{ij})] = \int_b \int_c \Phi(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + b_{ij})(\Delta_{ij} + c_{ij}) \phi^{(2)}(b_{ij}, c_{ij}; \mathbf{0}, \boldsymbol{\Sigma}_{ij}) db_{ij} dc_{ij}$$

where

$$\Sigma_{ij} = \begin{pmatrix} \sigma_{bij}^2 & \sigma_{bij}\sigma_{cij}\rho_{bcij} \\ \sigma_{bij}\sigma_{cij}\rho_{bcij} & \sigma_{cij}^2 \end{pmatrix}$$

and  $\sigma_{bij}^2$ ,  $\sigma_{cij}^2$  and  $\rho_{bcij}$  are the variances and correlation of  $B_i(t_{ij})$  and  $C_i(t_{ij})$ , respectively. Similarly, the overall marginal mean for a shared process model is given by

$$\mathbb{E}[g(Y_{ij})] = \int_b \Phi(\mathbf{X}_{ij}^\top \boldsymbol{\beta} + b_{ij})(\Delta_{ij} + \theta b_{ij})\phi^{(1)}(b_{ij}; 0, \sigma_{bij}^2)db_{ij}.$$

Conveniently, these integrals can be computed analytically and this results in

$$\Phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right) \Delta_{ij} + \frac{\sigma_{bij}\sigma_{cij}\rho_{bcij}}{\sqrt{1 + \sigma_{bij}^2}} \phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right)$$

and

$$\Phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right) \Delta_{ij} + \frac{\theta\sigma_{bij}^2}{\sqrt{1 + \sigma_{bij}^2}} \phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right)$$

respectively. The derivation of the first overall marginal mean of  $g(Y_{ij})$  (resulting from models with two processes) can be found in the supplementary material of Tom et al.,<sup>5</sup> and the second overall marginal mean of  $g(Y_{ij})$  is derived in the appendix. If we specify  $\mathbb{E}[g(Y_{ij})] = \mathbf{Z}_{ij}^\top \boldsymbol{\alpha}$ , we can then reparametrise

$$\Delta_{ij} = \left[ \mathbf{Z}_{ij}^\top \boldsymbol{\alpha} - \frac{\sigma_{bij}\sigma_{cij}\rho_{bcij}}{\sqrt{1 + \sigma_{bij}^2}} \phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right) \right] / \Phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right) \quad (22)$$

and

$$\Delta_{ij} = \left[ \mathbf{Z}_{ij}^\top \boldsymbol{\alpha} - \frac{\theta\sigma_{bij}^2}{\sqrt{1 + \sigma_{bij}^2}} \phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right) \right] / \Phi\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_{bij}^2}}\right)$$

in the respective models. Thus, as in equation (20),  $\boldsymbol{\alpha}$  offers easily interpretable covariate effects of  $\mathbf{Z}_{ij}$  on the overall marginal mean of  $g(Y_{ij})$  (by definition). In particular, a unit change in components of  $\mathbf{Z}_{ij}$  will increase the overall marginal mean of  $g(Y_{ij})$  by the respective components in  $\boldsymbol{\alpha}$ . However, from equations (21) and (22), it is also evident that replacing  $\mathbf{Z}_{ij}^\top \boldsymbol{\gamma}$  with  $\Delta_{ij}$  in Section 4 will still allow the proposed efficient estimation procedure to be applied. It is also possible to reparametrise patient-specific covariate effects  $\boldsymbol{\beta}$  in the binary component in terms of population-level covariate effects  $\boldsymbol{\xi}$ , specifically  $\boldsymbol{\beta} = \boldsymbol{\xi} \sqrt{1 + \sigma_{bij}^2}$ , since it can be shown that  $\mathbb{P}(U_{ij} = 1) = \Phi(\mathbf{X}_{ij}^\top \boldsymbol{\beta} / \sqrt{1 + \sigma_{bij}^2})$ . This relationship is easily proved. In the motivating application, this reparametrisation led to a numerically unstable optimization routine, therefore  $(\boldsymbol{\beta}, \boldsymbol{\alpha})$  was estimated with  $\hat{\boldsymbol{\xi}}$  obtained as  $\hat{\boldsymbol{\beta}} / \sqrt{1 + \hat{\sigma}_{bij}^2}$  and standard errors were calculated using the delta method.

## 6.1 Population-based inference

Using the parameterisations described in the previous subsection, we demonstrate how population-based inference on the probability of being disabled and the overall marginal mean HAQ score can be obtained. Specifically, on averaging across patients, how a unit change in covariate values impacts these quantities. For illustrative purposes,

the same covariates as those considered in the patient-specific case are considered. Note that for generalized linear models, conditional and marginal covariate effects will generally differ unless certain random effects distributions and link functions are chosen.<sup>25</sup>

As mentioned, marginal covariate effects on the probability of being disabled,  $\xi$ , were obtained from  $\beta/\sqrt{1 + \sigma_{bij}^2}$  with  $\beta$  and  $\sigma_{bij}^2$  (the variance of  $B_i(t_{ij})$ ) estimated using the model fitting procedure. The shared OU process and random effect models, where  $\sigma_{bij}^2$  does not depend on  $j$  (or  $i$ ) for these models, were considered. For models with random walks,  $\sigma_{bij}^2$  varies with  $j$  and therefore  $\xi$  will have a time-dependent interpretation. For simplicity, these models are not considered. The shared random effect model was fitted using the parameterisation described by equation (20), with  $g_1(y) = \Phi(y)$  and  $g_2(y) = y$ , and using the parameterisation described by equations (21) and (22), thus the same link functions are used and the inferences (at the population-level) from these models are comparable. These models will be denoted as shared random effect model-overall and -conditional respectively. Note that unlike at the population-level, the patient-specific assumptions from the shared random effect model-overall and -conditional are vastly different. The shared random effect-overall model assumes that the overall patient-specific mean, i.e.  $E[Y_{ij}|B_i^r = b_i^r, C_i^r = c_i^r]$ , has a linear form, namely  $Z_{ij}^T \alpha + c_i^r$ . While the shared random effect-conditional model assumes that this quantity takes a particular non-linear form, namely  $(\Delta_{ij} + c_i^r)\Phi(X_{ij}^T \beta + b_i^r)$ . As before, the shared random effect models (both -overall and -conditional) were fitted using numerical integration and maximum likelihood estimation under the assumption that the positive values follow a normal distribution with constant variance.

Table 2 presents the results. Population-level covariate effects on the overall marginal mean are seen to be relatively similar across models due to the considerable overlap in confidence intervals. All three models are in strong agreement regarding the population-level covariate effect of the number of active joints. That is, on average, patients with an additional actively inflamed joint have an overall mean HAQ score increased by approximately 0.02. In contrast to the patient-specific case, the population-level covariate effects on the probability of being disabled are now more consistent across models. A generalized likelihood ratio test of  $\rho_{gb} = 1$  produced a  $p$  value of  $< 0.001$  and therefore the shared OU process model is to be preferred over the shared random effect-conditional

**Table 2.** Table displaying population-level effects and corresponding 95% Wald intervals on the probability of being disabled and the overall marginal mean HAQ score.

|                                     | Shared OU process        | Shared random effect-conditional | Shared random effect-overall |
|-------------------------------------|--------------------------|----------------------------------|------------------------------|
| Binary component                    |                          |                                  |                              |
| Damaged joints                      | 0.012 (0.006, 0.018)     | 0.0057 (0.0006, 0.011)           | 0.0066 (0.002, 0.011)        |
| Active joints                       | 0.05 (0.043, 0.058)      | 0.063 (0.054, 0.071)             | 0.051 (0.044, 0.058)         |
| Sex                                 | -0.65 (-0.88, -0.43)     | -0.64 (-0.8, -0.48)              | -0.61 (-0.76, -0.46)         |
| Arthritis duration                  | 0.013 (0.0057, 0.02)     | 0.014 (0.0095, 0.02)             | 0.012 (0.0078, 0.017)        |
| Age at arthritis onset <sup>a</sup> | 0.19 (0.14, 0.24)        | 0.2 (0.13, 0.27)                 | 0.19 (0.12, 0.26)            |
| Intercept                           | 0.59 (0.35, 0.83)        | 0.56 (0.42, 0.7)                 | 0.68 (0.55, 0.81)            |
| Overall marginal mean               |                          |                                  |                              |
| Damaged Joints                      | 0.0064 (0.0036, 0.0094)  | 0.0029 (0.00054, 0.0053)         | 0.004 (0.0016, 0.0065)       |
| Active joints                       | 0.02 (0.018, 0.022)      | 0.021 (0.019, 0.023)             | 0.022 (0.02, 0.024)          |
| Sex                                 | -0.3 (-0.4, -0.19)       | -0.28 (-0.36, -0.21)             | -0.31 (-0.39, -0.23)         |
| Arthritis duration                  | 0.0035 (0.00032, 0.0068) | 0.006 (0.0038, 0.0082)           | 0.0054 (0.0031, 0.0077)      |
| Age at arthritis onset <sup>a</sup> | 0.073 (0.049, 0.098)     | 0.08 (0.048, 0.11)               | 0.09 (0.052, 0.13)           |
| Intercept                           | 0.62 (0.5, 0.73)         | 0.61 (0.54, 0.68)                | 0.59 (0.52, 0.66)            |
| $\theta$                            | 0.18 (0.16, 0.2)         | 0.27 (0.24, 0.29)                | 0.22 (0.2, 0.24)             |
| $\sigma^2$                          | 0.064 (0.057, 0.071)     | 0.12 (0.11, 0.12)                | 0.12 (0.11, 0.13)            |
| $\sigma_b^2$                        |                          | 3.63 (2.92, 4.5)                 | 5.19 (4.26, 6.34)            |
| $\sigma_{gb}^2$                     | 11.12 (8.71, 14.19)      |                                  |                              |
| $\rho_{gb}$                         | 0.95 (0.94, 0.96)        |                                  |                              |
| Log-likelihood                      | -3277.2                  | -3507.63                         | -3582.78                     |

<sup>a</sup>Denotes the standardised version of the covariate.

model. Log-likelihood values also indicate slight preference to the shared random effect-conditional model ( $-3507.63$ ) over the shared random effect-overall model ( $-3582.78$ ).

## 7 Discussion

This paper reconsiders the flexible two-part models of Albert and Shen<sup>8</sup> and Ghosh and Albert<sup>9</sup> and proposes an efficient method of implementation. Specifically, the problem of integrating over high dimensional random effects is replaced by evaluating the cumulative distribution function of a multivariate normal distribution. This leads to efficient algorithms being employed and results in only an optimization procedure being required for model fitting. Furthermore, while retaining the flexibility of including stochastic processes and the practicality of an efficient model fitting procedure, this paper also provides model parameterisations which allow easily interpretable covariate effects to act on the overall marginal mean. The proposed methodology was applied to a psoriatic arthritis data set with extensive follow-up information.

Through their application and a simulation study, Albert and Shen<sup>8</sup> demonstrated that overall conditional means (conditional on realisations of stochastic processes) may suffer from bias if serial correlation is present but a shared random effect model is used instead. Furthermore, as the shared random effect model becomes more misspecified ( $\rho_{gb}$  decreases from one), the degree of bias increases. However, under the same set-up, overall marginal means were less susceptible to bias. In the motivating application, the estimated degradation parameters from the shared OU process models were  $\hat{\rho}_{gb} = 0.95$  in both applications (Sections 5 and 6.1). The reasonably high estimated correlation may therefore explain why the shared random effect model ( $\rho_{gb} = 1$ ) was a reasonable approximation in terms of estimating regression coefficients, although it was substantially the worst fitting model.

Preliminary analyses suggested shared process models were reasonable for our data since  $\hat{\rho}$ ,  $\hat{\rho}_w$  and  $\hat{\rho}_g \approx 1$  when the described bivariate processes models were fitted. Although this may not be surprising as both parts of the model are describing the same response process, it is worth noting that the estimated correlation parameter (between processes) can in principle take a value between  $(-1, 1)$  as evidenced in other works.<sup>7,9</sup> Our preliminary analyses also demonstrated the need for careful evaluation of models fitted as problems with over fitting may arise. This was evident when the estimated random variation parameter was estimated to be virtually zero (i.e.  $\hat{\sigma}^2 \approx 0$ ) and the observed Fisher information matrix was non positive-definite, even when a considerably smaller tolerance level than the default was specified for computing multivariate normal probabilities.

As mentioned in Section 6, the proposed model parameterisations were motivated by making inference on the overall marginal mean. In this regard, covariate effects (both patient-specific and population-level) on the mean of the positive values and its correlation structure were assumed not of interest. If the mean of the positive values is of primary interest, it would be more sensible to directly use equations (18) and (19), as in Section 5, to obtain patient-specific effects or derive a similar parameterisation, as in Section 6, to obtain population-level effects.

A limitation of the current framework is that it is based on the assumption  $g(Y_{ij})|Y_{ij} > 0$  is approximately Gaussian with constant variance  $\sigma^2$ . Specifically, in situations where  $g(\cdot)$  is required to be complex so that this assumption will at least approximately hold, the resulting inferential targets will no longer be intuitively interpretable owing to the complexity of the transformation function. One approach that may weaken the need to assume normality of  $g(Y_{ij})|Y_{ij} > 0$ , particularly when the outcome exhibits a large amount of right skewness (e.g. medical expenditures), would be to make the alternative assumption  $g(Y_{ij})|Y_{ij} > 0$  follows a log-normal distribution. This may allow less complex and hence more interpretable transformation functions to be applied to the outcome without having to strongly violate the assumption on  $g(Y_{ij})|Y_{ij} > 0$ . Under this alternative assumption, we provide details in the supplementary materials of how easily interpretable inference on the overall marginal mean and on the mean of the positive transformed outcomes can be obtained with computationally efficient likelihoods. Similar techniques to those in the supplementary materials can also be used when the assumption  $g(Y_{ij})|Y_{ij} > 0$  follows a log-skew-normal distribution is of interest. Although this comes at the cost of having an increased number of integrations in the marginal likelihood.

Finally, the model described by equations (18) and (19) with possible simplifications described in Appendix 2 is very general. Although it was derived in the context of longitudinal semicontinuous data, it contains the model described by Barrett et al.<sup>15</sup> for the longitudinal and survival outcomes setting and implicitly provides a model for clustered cross-sectional semicontinuous data, where the index  $(i, j)$  specifies the  $j$ th outcome from the  $i$ th cluster. The multivariate normal identity described in Section 4.2 can also facilitate the fitting of flexible models describing clustered binary data and continuous bounded outcome data.<sup>14</sup> However, it should be noted that care is required when specifying an appropriate/suitable correlation structure. Particularly, the covariance



matrix must be constrained to be symmetric and positive semi-definite otherwise the model fitting procedure will likely be problematic, as was found here. For these alternative situations, the proposed methodology does nevertheless offer a strong basis, especially with regard to implementation, for the developing of flexible models.

## Acknowledgments

We are grateful to Professor Vernon T. Farewell for providing general discussions on this research. We also acknowledge the patients in the Toronto Psoriatic Arthritis Clinic.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was financially supported by the UK Medical Research Council [Unit program numbers U105261167 and MC\_UP\_1302/3].

## References

1. Olsen MK and Schaefer JL. A two-part random effects model for semicontinuous longitudinal data. *J Am Stat Assoc* 2001; **96**: 730–745.
2. Xie H, McHugo G, Sengupta A, et al. A method for analyzing longitudinal outcomes with many zeros. *Ment Health Serv Res* 2004; **6**: 239–246.
3. Smith VA, Preisser JS, Neelon B, et al. A marginalized two-part model for semicontinuous data. *Stat Med* 2014; **33**: 4891–4930.
4. Smith VA, Neelon B, Preisser JS, et al. A marginalized two-part model for longitudinal semicontinuous data. *Stat Meth Med Res* 2015; **July**: 1–24.
5. Tom BDM, Su L and Farewell VT. A corrected formulation for marginal inference derived from two-part mixed models for longitudinal semi-continuous data. *Stat Methods Med Res* 2016; **25**: 2014–2020.
6. Hall DB and Zhang Z. Marginal models for zero-inflated clustered data. *Stat Model* 2004; **4**: 161–180.
7. Tooze JA, Grunwald GK and Jones RH. Analysis of repeated measures data with clumping at zero. *Stat Meth Med Res* 2002; **11**: 341–355.
8. Albert PS and Shen J. Modelling longitudinal semicontinuous emesis volume data with serial correlation in an acupuncture clinical trial. *J R Stat Soc: Ser C* 2005; **54**: 707–720.
9. Ghosh P and Albert PS. A Bayesian analysis for longitudinal semicontinuous data with an application to an acupuncture clinical trial. *Comput Stat Data Anal* 2009; **53**: 699–706.
10. Bruce B and Fries JF. The Stanford health assessment questionnaire: dimensions and practical applications. *Health Qual Life Outcomes* 2003; **1**: 1–20.
11. Husted JA, Tom BD, Farewell VT, et al. A longitudinal study of the effect of disease activity and clinical damage on physical function over the course of psoriatic arthritis: does the effect change over time? *Arthritis Rheum* 2007; **56**: 840–849.
12. Su L, Tom BD and Farewell VT. Bias in two-part mixed models for longitudinal data. *Biostatistics* 2009; **10**: 374–389.
13. Su L, Tom BD and Farewell VT. A likelihood-based two-part marginal model for longitudinal semicontinuous data. *Stat Meth Med Res* 2015; **24**: 194–205.
14. Huttmacher MM, French JL, Krishnaswami S, et al. Estimating transformations for repeated measures modelling of continuous bounded outcome data. *Stat Med* 2010; **30**: 935–949.
15. Barrett J, Diggle P, Henderson R and Taylor-Robinson D. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *J R Stat Soc: Ser B* 2015; **77**: 131–148.
16. Azzalini A. A class of distributions which includes the normal ones. *Scand J Stat* 1985; **12**: 171–178.
17. Azzalini A. The skew-normal distribution and related multivariate families (with discussion). *Scand J Stat* 1985; **32**: 159–200.
18. Arnold BC and Beaver RJ. Hidden truncation models. *Sankhya A* 2000; **62**: 22–35.
19. Arnold BC. Flexible univariate and multivariate models based on hidden truncation. *J Stat Plan* 2009; **139**: 3741–3749.
20. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria, <http://www.R-project.org>, ISBN 3-900051-07-0. (accessed 6 May 2017).



21. Azzalini A. *mnormt: the multivariate normal and t distributions*. R package version 1.5-2, <http://azzalini.stat.unipd.it/SW/Pkg-mnormt> (2015).
22. Broyden CG. The convergence of a class of double-rank minimisation algorithms. *J Inst Math Appl* 1970; **6**: 76–90.
23. Henderson R and Shimakura S. A serially correlated gamma frailty model for longitudinal count data. *Biometrika* 2003; **90**: 355–366.
24. Albert PS. Letter to the editor. *Biometrics* 2005; **47**: 879–881.
25. Diggle P, Heagerty P, Liang KY, et al. *Analysis of longitudinal data*. New York: Oxford University Press, 2002.
26. Owen DB. A table of normal integrals. *Commun Stat B Simul* 1980; **9**: 389–419.

## Appendix 1. Rearranging equation (11)

The continuous and random effects component in the integrand of equation (9) is

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\sum_{j=1}^{m_i} u_{ij}} \exp\left\{-\frac{(g(y_i) - A_{i3}\mu_{ic} - A_{i4}l_i)^\top (g(y_i) - A_{i3}\mu_{ic} - A_{i4}l_i)}{2\sigma^2}\right\} \phi^{\{\dim(l_i)\}}(l_i; \mathbf{0}, \Sigma_{il}).$$

We rearrange this expression by completing the square in  $l_i$ . This results in

$$\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\sum_{j=1}^{m_i} u_{ij}} \exp\left\{-\frac{(g(y_i) - A_{i3}\mu_{ic})^\top (g(y_i) - A_{i3}\mu_{ic})}{2\sigma^2}\right\} \frac{|H_i^{-1}|^{1/2}}{|\Sigma_{il}|^{1/2}} \frac{1}{(2\pi)^{\dim(l_i)/2} |H_i^{-1}|^{1/2}} \exp\left(-\frac{l_i^\top H_i l_i}{2} + l_i^\top H_i h_i\right). \quad (23)$$

Focusing on terms containing  $l_i$ , we have

$$\exp\left(-\frac{l_i^\top H_i l_i}{2} + l_i^\top H_i h_i\right) = \exp\left\{-\frac{(l_i - h_i)^\top H_i (l_i - h_i)}{2} + \frac{h_i^\top H_i h_i}{2}\right\}. \quad (24)$$

Equations (12) to (14) now follows by substituting equation (24) into equation (23).

## Appendix 2. Simplification for correlated stochastic processes model

For the models containing correlated stochastic processes (described by equation (7)), recall that  $\Sigma_{il} = \Sigma_i$  (where  $\Sigma_i$  is described by either equation (5) or equation (6)),  $A_{i2} = (\text{diag}(2\mathbf{u}_i - \mathbf{1}), \text{diag}(\mathbf{0}))$  and  $A_{i4} = (\text{diag}(\mathbf{0}), \mathbf{u}_i)$  are  $m_i \times 2m_i$  matrices. Simplification of equation (18) for this model structure is possible, specifically the following (to be derived) equations

$$\begin{aligned} A_{i2}h_i &= (\Sigma_{ibc}^u - \Sigma_{ibc}^u(I_{m_i} + \Sigma_{ic}^u/\sigma^2)^{-1}\Sigma_{ic}^u/\sigma^2)(g(y_i) - A_{i3}\mu_{ic})/\sigma^2 \\ A_{i2}H_i^{-1}A_{i2}^\top &= \Sigma_{ib}^u - \Sigma_{ibc}^u(I_{m_i} + \Sigma_{ic}^u/\sigma^2)^{-1}\Sigma_{ibc}^u/\sigma^2 \\ |\Sigma_i H_i| &= |\Sigma_{ic} \text{diag}(\mathbf{u}_i)/\sigma^2 + I_{m_i}| \\ h_i^\top H_i h_i &= (g(y_i) - A_{i3}\mu_{ic})^\top (\Sigma_{ic}^u - \Sigma_{ic}^u(I_{m_i} + \Sigma_{ic}^u/\sigma^2)^{-1}\Sigma_{ic}^u/\sigma^2)(g(y_i) - A_{i3}\mu_{ic})/\sigma^4 \end{aligned} \quad (25)$$

where

$$\begin{aligned} \Sigma_{ib}^u &= \text{diag}(2\mathbf{u}_i - \mathbf{1})\Sigma_{ib}\text{diag}(2\mathbf{u}_i - \mathbf{1}) \\ \Sigma_{ic}^u &= \text{diag}(\mathbf{u}_i)\Sigma_{ic}\text{diag}(\mathbf{u}_i) \\ \Sigma_{ibc}^u &= \text{diag}(2\mathbf{u}_i - \mathbf{1})\Sigma_{ibc}\text{diag}(\mathbf{u}_i) \end{aligned}$$

reduce the dimension of the respective matrix calculations (from  $2m_i$  to  $m_i$  dimensional).

In order to derive the equations in (25), we begin by simplifying  $\Sigma_i H_i$ . That is

$$\Sigma_i H_i = \Sigma_i A_{i4}^\top A_{i4}/\sigma^2 + I_{2m_i} = \begin{pmatrix} \Sigma_{ib} & \Sigma_{ibc} \\ \Sigma_{ibc} & \Sigma_{ic} \end{pmatrix} \begin{pmatrix} \text{diag}(\mathbf{0}) & \text{diag}(\mathbf{0}) \\ \text{diag}(\mathbf{0}) & \text{diag}(\mathbf{u}_i)/\sigma^2 \end{pmatrix} + I_{2m_i}$$

$$= \begin{pmatrix} \mathbf{I}_{m_i} & \boldsymbol{\Sigma}_{ibc} \text{diag}(\mathbf{u}_i)/\sigma^2 \\ \text{diag}(\mathbf{0}) & \boldsymbol{\Sigma}_{ic} \text{diag}(\mathbf{u}_i)/\sigma^2 + \mathbf{I}_{m_i} \end{pmatrix}.$$

It now follows that

$$|\boldsymbol{\Sigma}_i \mathbf{H}_i| = |\mathbf{I}_{m_i}| |\boldsymbol{\Sigma}_{ic} \text{diag}(\mathbf{u}_i)/\sigma^2 + \mathbf{I}_{m_i}| = |\boldsymbol{\Sigma}_{ic} \text{diag}(\mathbf{u}_i)/\sigma^2 + \mathbf{I}_{m_i}|.$$

Next, we simplify  $\mathbf{h}_i^\top \mathbf{H}_i \mathbf{h}_i = (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})^\top \mathbf{A}_{i4} \mathbf{H}_i^{-1} \mathbf{A}_{i4}^\top (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})/\sigma^4$ . By using the Woodbury matrix identity, specifically

$$\mathbf{H}_i^{-1} = \boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_i \mathbf{A}_{i4}^\top (\mathbf{I}_{m_i} + \mathbf{A}_{i4} \boldsymbol{\Sigma}_i \mathbf{A}_{i4}^\top / \sigma^2)^{-1} \mathbf{A}_{i4} \boldsymbol{\Sigma}_i / \sigma^2$$

and noting that

$$\mathbf{A}_{i4} \boldsymbol{\Sigma}_i \mathbf{A}_{i4}^\top = \text{diag}(\mathbf{u}_i) \boldsymbol{\Sigma}_{ic} \text{diag}(\mathbf{u}_i) \equiv \boldsymbol{\Sigma}_{ic}^u$$

we have

$$\begin{aligned} \mathbf{h}_i^\top \mathbf{H}_i \mathbf{h}_i &= (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})^\top \mathbf{A}_{i4} \mathbf{H}_i^{-1} \mathbf{A}_{i4}^\top (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})/\sigma^4 \\ &= (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})^\top (\boldsymbol{\Sigma}_{ic}^u - \boldsymbol{\Sigma}_{ic}^u (\mathbf{I}_{m_i} + \boldsymbol{\Sigma}_{ic}^u / \sigma^2)^{-1} \boldsymbol{\Sigma}_{ic}^u / \sigma^2) (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})/\sigma^4. \end{aligned}$$

To simplify  $\mathbf{A}_{i2} \mathbf{h}_i$ , consider

$$\begin{aligned} \mathbf{A}_{i2} \mathbf{h}_i &= \mathbf{A}_{i2} \mathbf{H}_i^{-1} \mathbf{A}_{i4}^\top (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})/\sigma^2 \\ &= (\mathbf{A}_{i2} \boldsymbol{\Sigma}_i \mathbf{A}_{i4}^\top - \mathbf{A}_{i2} \boldsymbol{\Sigma}_i \mathbf{A}_{i4}^\top (\mathbf{I}_{m_i} + \boldsymbol{\Sigma}_{ic}^u / \sigma^2)^{-1} \boldsymbol{\Sigma}_{ic}^u / \sigma^2) (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})/\sigma^2. \end{aligned}$$

By noting that

$$\mathbf{A}_{i2} \boldsymbol{\Sigma}_i \mathbf{A}_{i4}^\top = \text{diag}(2\mathbf{u}_i - \mathbf{1}) \boldsymbol{\Sigma}_{ibc} \text{diag}(\mathbf{u}_i) \equiv \boldsymbol{\Sigma}_{ibc}^u$$

we have

$$\mathbf{A}_{i2} \mathbf{h}_i = (\boldsymbol{\Sigma}_{ibc}^u - \boldsymbol{\Sigma}_{ibc}^u (\mathbf{I}_{m_i} + \boldsymbol{\Sigma}_{ic}^u / \sigma^2)^{-1} \boldsymbol{\Sigma}_{ic}^u / \sigma^2) (g(\mathbf{y}_i) - \mathbf{A}_{i3} \boldsymbol{\mu}_{ic})/\sigma^2.$$

Furthermore

$$\mathbf{A}_{i2} \mathbf{H}_i^{-1} \mathbf{A}_{i2}^\top = \boldsymbol{\Sigma}_{ib}^u - \boldsymbol{\Sigma}_{ibc}^u (\mathbf{I}_{m_i} + \boldsymbol{\Sigma}_{ic}^u / \sigma^2)^{-1} \boldsymbol{\Sigma}_{ibc}^u / \sigma^2$$

where

$$\boldsymbol{\Sigma}_{ib}^u \equiv \mathbf{A}_{i2} \boldsymbol{\Sigma}_i \mathbf{A}_{i2}^\top = \text{diag}(2\mathbf{u}_i - \mathbf{1}) \boldsymbol{\Sigma}_{ib} \text{diag}(2\mathbf{u}_i - \mathbf{1}).$$

### Appendix 3. Overall marginal mean of shared process model

To obtain the overall marginal mean of a shared process model, the following integral must be evaluated

$$\begin{aligned} \int_b \Phi(\mathbf{X}^\top \boldsymbol{\beta} + b) (\mathbf{Z}^\top \boldsymbol{\gamma} + \theta b) \phi^{(1)}(b; 0, \sigma_b^2) db &= \mathbf{Z}^\top \boldsymbol{\gamma} \int_b \Phi(\mathbf{X}^\top \boldsymbol{\beta} + b) \phi^{(1)}(b; 0, \sigma_b^2) db \\ &\quad + \theta \int_b b \Phi(\mathbf{X}^\top \boldsymbol{\beta} + b) \phi^{(1)}(b; 0, \sigma_b^2) db. \end{aligned} \tag{26}$$

The first term in equation (26) can be evaluated using the skew normal result, i.e. using equation (10), that is

$$\mathbf{Z}^\top \gamma \int_b \Phi(\mathbf{X}^\top \boldsymbol{\beta} + b) \phi^{(1)}(b; 0, \sigma_b^2) db = \mathbf{Z}^\top \gamma \Phi^{(1)}(\mathbf{X}^\top \boldsymbol{\beta}; 0, 1 + \sigma_b^2) = \mathbf{Z}^\top \gamma \Phi\left(\frac{\mathbf{X}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_b^2}}\right).$$

To compute the second term in equation (26), consider

$$\theta \int_b b \Phi(\mathbf{X}^\top \boldsymbol{\beta} + b) \phi^{(1)}(b; 0, \sigma_b^2) db = \theta \sigma_b \int_{b^*} b^* \Phi(\mathbf{X}^\top \boldsymbol{\beta} + \sigma_b b^*) \phi(b^*) db^*.$$

This integral can be computed using equation (10) (011.3) in Owen,<sup>26</sup> which results in

$$\frac{\theta \sigma_b^2}{\sqrt{1 + \sigma_b^2}} \phi\left(\frac{\mathbf{X}^\top \boldsymbol{\beta}}{\sqrt{1 + \sigma_b^2}}\right).$$

The overall marginal mean now follows.